

(12) DEMANDE INTERNATIONALE PUBLIÉE EN VERTU DU TRAITÉ DE COOPÉRATION
EN MATIÈRE DE BREVETS (PCT)

(19) Organisation Mondiale de la Propriété
Intellectuelle
Bureau international



1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100

(43) Date de la publication internationale
7 juin 2001 (07.06.2001)

PCT

(10) Numéro de publication internationale
WO 01/41125 A1

(51) Classification internationale des brevets⁷: G10L 15/18

(72) Inventeurs; et

(21) Numéro de la demande internationale:

PCT/FR00/03329

(75) Inventeurs/Déposants (pour US seulement): SOUF-
FLET, Frédéric [FR/FR]; Thomson Multimedia, 46,
quai Alphonse Le Gallo, F-92648 Boulogne (FR). DE-
LAUNAY, Christophe [FR/FR]; Thomson Multimedia,
46, quai Alphonse Le Gallo, F-92648 Boulogne (FR).
TAZINE, Nour-Eddine [FR/FR]; Thomson Multimedia,
46, quai Alphonse Le Gallo, F-92648 Boulogne (FR).

(22) Date de dépôt international:

29 novembre 2000 (29.11.2000)

(25) Langue de dépôt:

français

(74) Mandataire: KOHRS, Martin; Thomson Multimedia,
46, quai Alphonse Le Gallo, F-92648 Boulogne (FR).

(26) Langue de publication:

français

(30) Données relatives à la priorité:

99/15190 2 décembre 1999 (02.12.1999) FR

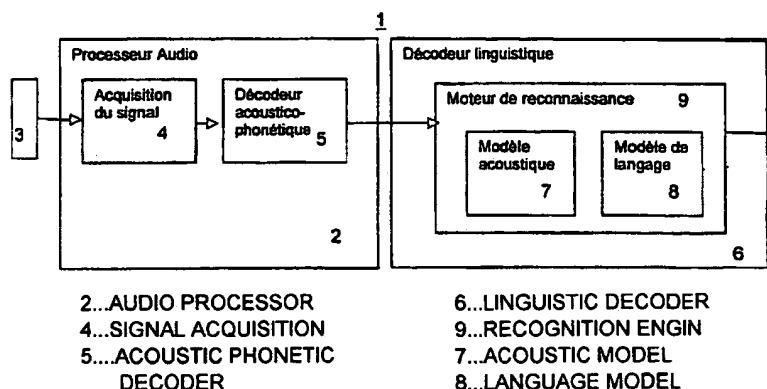
(81) États désignés (national): AE, AG, AL, AM, AT, AU, AZ,
BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE,
DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU,
ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS,
LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO,
NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR,
TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(71) Déposant (pour tous les États désignés sauf US): THOM-
SON MULTIMEDIA [FR/FR]; 46, quai Alphonse Le
Gallo, F-92100 Boulogne-Billancourt (FR).

[Suite sur la page suivante]

(54) Title: SPEECH RECOGNITION WITH A COMPLEMENTARY LANGUAGE MODEL FOR TYPICAL MISTAKES IN
SPOKEN DIALOGUE

(54) Titre: RECONNAISSANCE DE PAROLE AVEC UN MODELE DE LAGAGE COMPLEMENTAIRE POUR LES ERREURS
TYPES DU DIALOGUE PARLE



(57) Abstract: The invention concerns a speech recognition device (1) comprising an audio processor (2) for acquiring an audio signal and a linguistic decoder (6) for determining a sequence of words corresponding to the audio signal. The linguistic decoder of the inventive device comprises a language model (8) determined from a first set of at least a syntactic block defined only by a grammar and a second set of at least a second syntactic block defined by one of the following elements, or a combination of said elements: a grammar, a list of sentence segments, a n-gram system.

(57) Abrégé: L'invention concerne un dispositif de reconnaissance vocale (1) comportant un processeur audio (2) pour l'acquisition d'un signal audio et un décodeur linguistique (6) pour déterminer une séquence de mots correspondant au signal audio. Le décodeur linguistique du dispositif de l'invention comporte un modèle de langage (8) déterminé à partir d'un premier ensemble d'au moins un bloc syntaxique défini uniquement par une grammaire et d'un second ensemble d'au moins un second bloc syntaxique défini par l'un des éléments suivants, ou une combinaison de ces éléments: une grammaire, une liste de tronçons de phrase, un réseau n-gram.

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100

WO 01/41125 A1



(84) États désignés (régional): brevet ARIPO (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), brevet eurasien (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), brevet européen (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), brevet OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

En ce qui concerne les codes à deux lettres et autres abréviations, se référer aux "Notes explicatives relatives aux codes et abréviations" figurant au début de chaque numéro ordinaire de la Gazette du PCT.

Publiée:

— Avec rapport de recherche internationale.

RECONNAISSANCE DE PAROLE AVEC UN MODELE DE LANGAGE COMPLEMENTAIRE POUR LES ERREURS
TYPES DU DIALOGUE PARLE

5 L'invention concerne un dispositif de reconnaissance vocale comportant un modèle de langage défini à l'aide de blocs syntaxiques de différente nature, dits blocs rigides et blocs souples.

10 Les systèmes d'information ou de contrôle utilisent de plus en plus souvent une interface vocale pour rendre l'interaction avec l'utilisateur rapide et intuitive. Ces systèmes devenant plus complexes, les styles de dialogue supportés sont de plus en plus riches, et l'on rentre dans le domaine de la reconnaissance vocale continue à très large vocabulaire.

15 On sait que la conception d'un système de reconnaissance vocale continue à large vocabulaire suppose de produire un *Modèle de Langage* qui définit la probabilité pour qu'un mot donné du vocabulaire de l'application suive un autre mot ou groupe de mots, dans l'ordre chronologique de la phrase.

20

Ce modèle de langage doit reproduire le style d'élocution ordinairement employé par un utilisateur du système : hésitations, faux départs, changements d'avis, etc.

25 La qualité du modèle de langage utilisé influe de manière forte sur la fiabilité de la reconnaissance vocale. Cette qualité est le plus souvent mesurée par un indice appelé perplexité du modèle de langage, et qui représente schématiquement le nombre de choix que le système doit réaliser pour chaque mot décodé. Plus cette perplexité est basse, plus la qualité est bonne.

30

Le modèle de langage est nécessaire pour traduire le signal vocal en une suite textuelle de mots, étape souvent utilisée par les systèmes de

dialogue. Il faut alors construire une logique de compréhension qui permette de comprendre la requête formulée vocalement pour y répondre.

Il existe deux méthodes standards pour produire des modèles de langage à large vocabulaire :

(1) La méthode statistique dite en N-gram, le plus souvent en bigram ou trigram, consiste à supposer que la probabilité d'occurrence d'un mot dans la phrase dépend uniquement des N mots qui le précèdent, indépendamment de son contexte dans la phrase.

Si l'on prend l'exemple du trigram pour un vocabulaire de 1000 mots, comme il existe 1000^3 groupes possibles de trois éléments, il faudrait définir 1000^3 probabilités pour définir le modèle de langage, ce qui occupe une taille mémoire considérable et une très grande puissance de calcul. Pour résoudre ce problème, les mots sont groupés en ensembles qui sont soit définis explicitement par le concepteur du modèle, soit déduits par des méthodes auto-organisatrices.

Ce modèle de langage est construit à partir d'un corpus de texte de façon automatique.

20

(2) La seconde méthode consiste à décrire la syntaxe au moyen d'une grammaire probabiliste, typiquement une grammaire non contextuelle définie grâce à un ensemble de règles décrites sous la forme dite Backus Naur Form ou BNF.

25

Les règles décrivant les grammaires sont le plus souvent écrites à la main, mais peuvent également être déduites de manière automatique. A ce titre, on peut se référer au document suivant :

"Basic methods of probabilistic context free grammars" par F. Jelinek, J.D. Lafferty et R.L. Mercer NATO ASI Series Vol. 75 pp. 345-359, 1992.

30

Les modèles précédemment décrits posent des problèmes spécifiques lorsqu'ils sont appliqués à des interfaces de systèmes en langage naturel :

5 Les modèles de langages de type N-grams (1) ne modélisent pas correctement les dépendances entre plusieurs sous-structures grammaticales distantes dans la phrase. Pour une phrase prononcée syntaxiquement correcte, rien ne garantit que ces sous-structures seront respectées au cours de la reconnaissance, ce qui fait qu'il est difficile de
10 déterminer si tel ou tel sens, habituellement porté par une ou plusieurs structures syntaxiques précises, est véhiculé par la phrase.

Ces modèles sont adaptés à la dictée continue, mais leur application dans les systèmes de dialogue souffre des défauts mentionnés.

Par contre, il est possible, dans un modèle de type N-gram, de
15 prendre en compte les hésitations et des répétitions, en définissant des ensembles de mots regroupant les mots qui ont effectivement été récemment prononcés.

Les modèles fondés sur des grammaires (2) permettent de
20 modéliser correctement les dépendances à distance dans une phrase, et également de respecter des sous-structures syntaxiques précises. La perplexité du langage obtenu est souvent plus faible, pour une application donnée, que pour les modèles de type N-gram.

Par contre, ils conviennent mal à la description d'un style de
25 langage parlé, avec prise en compte d'hésitations, de faux départs, etc. En effet, ces phénomènes liés au langage parlé ne peuvent se prédire, il paraît donc difficile de concevoir des grammaires qui, de par leur nature, sont basées sur des règles de langage.

De plus, le nombre de règles nécessaires pour couvrir une
30 application est très grand, ce qui rend difficile la prise en compte de nouvelles phrases à ajouter au dialogue prévu sans modification des règles existantes.

L'invention a pour objet un dispositif de reconnaissance vocale comportant un processeur audio pour l'acquisition d'un signal audio et un
5 décodeur linguistique pour déterminer une séquence de mots correspondant au signal audio, le décodeur comportant un modèle de langage (8)

caractérisé en ce que le modèle de langage (8) est déterminé par deux ensembles de blocs. Le premier ensemble comprend au moins un bloc syntaxique rigide et le second ensemble comprend au moins un bloc
10 syntaxique souple.

L'association des deux types de blocs syntaxiques permet de résoudre facilement les problèmes liés au langage parlé tout en bénéficiant de la modélisation des dépendances entre les éléments d'une phrase, modélisation facilement traitable à l'aide d'un bloc syntaxique rigide.

15 Selon une particularité, le premier ensemble de blocs syntaxiques rigides est défini par une grammaire de type BNF.

Selon une autre particularité, le second ensemble de blocs syntaxiques souples est défini par un ou plusieurs réseaux n-gram, les données des réseaux n-gram étant produites à l'aide d'une grammaire ou
20 d'une liste de tronçons de phrase.

Selon une autre particularité, les réseaux n-gram contenus dans les seconds blocs souples contiennent des données permettant de reconnaître les phénomènes de langage parlé suivants : l'hésitation simple, la répétition simple, l'échange simple, le changement d'avis, le bafouillage.

25 Le modèle de langage selon l'invention autorise la combinaison des avantages des deux systèmes, en définissant deux types d'entités qui se combinent pour former le modèle de langage final.

On conserve pour certaines entités une syntaxe rigide et on leur
30 associe un parseur, tandis que d'autres sont décrites par un réseau de type n-gram.

De plus, selon une variante de réalisation, on définit des blocs libres « déclenchés » par des blocs d'un des types précédents.

D'autres caractéristiques et avantages de l'invention apparaîtront à travers la description d'un mode de réalisation particulier non limitatif, explicité à l'aide des dessins joints parmi lesquels :

- la figure 1 est un diagramme d'un système de reconnaissance vocale,
- 10 - la figure 2 est un diagramme OMT définissant un bloc syntaxique selon l'invention.

La figure 1 est un diagramme bloc d'un exemple de dispositif 1 de reconnaissance de la parole. Ce dispositif comporte un processeur 2 du signal audio réalisant la numérisation d'un signal audio provenant d'un microphone 3 par l'intermédiaire d'un circuit 4 d'acquisition du signal. Le processeur traduit également les échantillons numériques en symboles acoustiques choisis dans un alphabet prédéterminé. Il comporte à cet effet un décodeur acoustico-phonétique 5. Un décodeur linguistique 6 traite ces symboles dans le but de déterminer, pour une séquence A de symboles, la séquence W de mots la plus probable, étant donné la séquence A.

Le décodeur linguistique utilise un modèle acoustique 7 et un modèle de langage 8 mis en œuvre par un algorithme de recherche par hypothèse 9. Le modèle acoustique est par exemple un modèle dit 'Markov caché' ('Hidden Markov model' ou HMM). Le modèle de langage mis en œuvre dans le présent exemple de réalisation est basé sur une grammaire décrite à l'aide de règles de syntaxe de forme Backus Naur. Le modèle de langage est utilisé pour soumettre des hypothèses à l'algorithme de recherche. Ce dernier, qui est le moteur de reconnaissance proprement dit, est pour ce qui du présent exemple un algorithme de recherche basé sur un algorithme de type Viterbi et appelé 'n-best'. L'algorithme de type n-best détermine à chaque étape de l'analyse d'une phrase les n séquences de

mots les plus probables. En fin de phrase, la solution la plus probable est choisie parmi les n candidats.

Les notions du paragraphe ci-dessus sont en soi bien connues de l'Homme du Métier, mais des informations concernant en particulier l'algorithme n-best sont donnés dans l'ouvrage :

"Statistical methods for speech recognition" par F. Jelinek, MIT Press 1999 ISBN 0-262-10066-5 pp. 79-84. D'autres algorithmes peuvent également être mis en œuvre. Notamment, d'autres algorithmes de type 'Beam Search' (Recherche par faisceau), dont l'algorithme 'n-best' est un exemple.

Le modèle de langage de l'invention utilise des blocs syntaxiques qui peuvent être de l'un des deux types illustrés par la figure 2: bloc de type rigide, bloc de type souple.

Les blocs syntaxiques rigides sont définis grâce à une syntaxe de type BNF, avec cinq règles d'écriture :

- (a) <symbole A> = <symbole B> | <symbole C> (symbole ou)
- (b) <symbole A> = <symbole B> <symbole C> (symbole et)
- (c) <symbole A> = <symbole B>? (symbole facultatif)
- (d) <symbole A> = "mot lexical" (affectation lexicale)
- (e) <symbole A> = P{<symbole B>, <symbole C> ,...<symbole X>}
(symbole B > <symbole C>)
(.....)
(symbole I > <symbole J>)

(toutes les permutations sans répétition des symboles cités, avec contraintes : le symbole B doit apparaître avant le symbole C, le symbole I avant le symbole J ...)

5 La mise en œuvre de la règle (e) est expliquée plus en détail dans la demande de brevet français n° 9915083 s'intitulant 'Dispositif de reconnaissance vocale mettant en œuvre une règle syntaxique de permutation' déposée au nom de THOMSON multimedia le 30 novembre 1999.

10

Les blocs souples sont définis soit grâce à la même syntaxe BNF que précédemment, soit comme une liste de tronçons de phrase, soit par une liste de vocabulaire et les réseaux n-gram correspondants, soit par la combinaison des trois. Mais ces informations sont traduites
15 systématiquement en un réseau n-gram et, si la définition a été faite par un fichier BNF, il n'est pas garanti que seules les phrases syntaxiquement correctes vis-à-vis de cette grammaire puissent être produites.

Un bloc souple est donc défini par une probabilité $P(S)$
20 d'apparition de la suite S de n mots w_i de la forme (dans le cas d'un trigram) :

$$P(S) = \prod_{i=1,n} P(w_i)$$

$$\text{Avec } P(w_i) = P(w_i | w_{i-1}, w_{i-2})$$

25 Pour chaque bloc souple, il existe un mot spécial de sortie de bloc qui apparaît dans le réseau n-gram de la même façon qu'un mot normal, mais qui n'a pas de trace phonétique et qui autorise à sortir du bloc.

Une fois que ces bloc syntaxiques ont été définis (de type n-gram
30 ou de type BNF), ils peuvent à nouveau être utilisés comme atomes pour des constructions d'ordre supérieur :

Dans le cas d'un bloc BNF, les blocs de niveaux inférieurs peuvent être utilisés en lieu et place de l'affectation lexicale ainsi que dans les autres règles.

5 Dans le cas d'un bloc de type n-gram, les blocs de niveau inférieur sont utilisés en lieu et place des mots w_i , et donc plusieurs blocs peuvent s'enchaîner avec une probabilité donnée.

10 Une fois le réseau n-gram défini, il est incorporé dans la grammaire BNF précédemment décrite comme un symbole particulier. On peut incorporer autant de réseaux n-gram que nécessaire dans la grammaire BNF. Les permutations utilisées pour la définition d'un bloc de type BNF sont traitées dans l'algorithme de recherche du moteur de reconnaissance par des variables de type booléen utilisées pour diriger la recherche lors de l'élagage classiquement mis en œuvre dans ce type de situation.

15

On voit que le symbole de sortie de bloc souple peut aussi être interprété comme un symbole de remontée au bloc supérieur, qui peut être lui-même un bloc souple ou un bloc rigide.

20 ○ **Mise en place de déclencheurs**

Le formalisme précédent n'est encore pas suffisant pour décrire le modèle de langage d'une application de dialogue homme-machine à large vocabulaire. Selon une variante de réalisation, on y ajoute un mécanisme de déclenchement ou 'trigger'. Le déclencheur permet de donner du sens à un mot ou à un bloc, afin de l'associer à certains éléments. Par exemple, 25 supposons que le mot « documentaire » est reconnu dans le contexte d'un guide électronique de programmes audiovisuels. On peut associer à ce mot une liste de mots tels que « animalier, sportif, touristique,... ». Ces mots ont 30 un sens pour « documentaire », et on peut s'attendre à ce que l'un d'entre eux soit associé à celui-ci.

Pour cela, nous noterons <bloc> un bloc précédemment décrit et ::<bloc> la réalisation de ce bloc par une de ses instances au cours de l'algorithme de reconnaissance, c'est-à-dire sa présence dans la chaîne actuellement décodée dans l'algorithme n-best search de recherche.

5

Par exemple, on pourrait avoir :

<souhait> = j'aimerais aller à | je veux me rendre à.

<ville> = Lyon | Paris | Londres | Rennes.

10

<phrase> = <souhait> <ville>

Alors ::<souhait> sera: 'j'aimerais aller à' pour la partie des chemins envisagé par l'algorithme de Viterbi pour les possibilités :

15

j'aimerais aller à Lyon

j'aimerais aller à Paris

j'aimerais aller à Londres

j'aimerais aller à Rennes

20

et sera égal à 'je veux me rendre à' pour les autres.

Les déclencheurs du modèle de langage sont alors définis de la manière suivante :

25

Si <symbole>:: appartient à un sous groupe donné des réalisations possibles du symbole en question, alors un autre symbole <T(symbole)>, qui est le symbole cible du symbole actuel; est soit réduit à une sous partie de son domaine d'extension normal, c'est-à-dire à son domaine d'extension si le déclencheur n'est pas présent dans la chaîne de

30 décodage, (déclencheur réducteur), soit activé et disponible, avec un facteur de branchement non nul en sortie de chaque bloc syntaxique appartenant au groupe dit des 'candidats activeurs' (déclencheur activateur)

Notons que :

Il n'est pas nécessaire que tous les blocs décrivent un processus
5 de déclenchement.

La cible d'une symbole peut être ce symbole lui-même, s'il est
utilisé de manière multiple dans le modèle de langage.

Il peut n'exister, pour un bloc, qu'une sous partie de son ensemble
de réalisation qui soit composante d'un mécanisme de déclenchement, le
10 complémentaire n'étant pas lui-même un déclencheur.

La cible d'un déclencheur activateur peut être un symbole facultatif.

Les mécanismes de déclenchement réducteurs permettent de
traiter, dans notre modèle de langage de blocs, les répétitions cohérentes de
15 thèmes. Des informations supplémentaires sur la notion de déclencheur
peuvent être trouvées dans le document de référence déjà cité, notamment
pages 245-253.

Les mécanismes de déclenchement activateurs permettent de
20 modéliser certains groupes syntaxiques libres, dans les langages fortement
infléchis.

Il faut noter que les déclencheurs, leurs cibles et la restriction sur
les cibles, peuvent être déterminés manuellement ou obtenus par un
25 processus automatique, par exemple par une méthode de maximum
d'entropie.

◦ **Prise en compte du langage parlé :**

30

La construction précédemment décrite définit la syntaxe du
modèle de langage, sans prise en compte des hésitations, reprises, faux

départs, changement d'avis, etc., qui sont attendus dans un style parlé. Les phénomènes liés au langage parlé sont difficilement reconnaissables par une grammaire, du fait de leur nature non-prédictible. Les réseaux n-gram sont plus adaptés pour reconnaître ce genre de phénomène.

5

Ces phénomènes liés au langage parlé peuvent être classés en cinq catégories :

L'hésitation simple : je voudrais (*heuuuu...silence*) aller à Lyon.

10

La répétition simple, dans laquelle une partie de la phrase, (souvent les déterminants et les articles, mais parfois des morceaux entiers de phrase), sont purement et simplement répétés : je voudrais aller à (à à à) Lyon.

L'échange simple, au cours de laquelle une formulation est remplacée, en cours de route, par une formulation de même sens, mais syntaxiquement différente : je voudrais me rendre (*heuuuu aller*) à Lyon

15

Le changement d'avis : une partie de phrase est corrigée, avec un sens différent, au cours de l'énoncé : je voudrais aller à Lyon, (*heuuuu à Paris*).

20

Le bafouillage : je voudrais aller à (*Praris Heuuu*) Paris.

Les deux premiers phénomènes sont les plus fréquents : environ 80% des hésitations sont classées dans l'un de ces groupes.

25

Le modèle de langage de l'invention traite ces phénomènes de la manière suivante :

Hésitation simple :

30

L'hésitation simple est traitée par création de mots associés aux traces phonétiques marquant l'hésitation dans la langue considérée, et qui

sont traités de la même façon que les autres vis-à-vis du modèle de langage (probabilité d'apparition, d'être suivi d'un silence, etc.), et dans les modèles phonétiques (coarticulation, etc.).

On a remarqué que des hésitations simples se produisent à des
5 endroits précis d'une phrase, par exemple : entre le premier verbe et le second verbe. Pour les traiter, un exemple de règle d'écriture conforme à la présente invention consiste :

<Groupe verbal> = <premier verbe><réseau n-gram><second verbe>

10

Répétition simple :

La répétition simple est traitée par une technique de cache qui
contient la phrase analysée actuellement à cette étape du décodage. Il
15 existe, dans le modèle de langage, une probabilité fixe pour qu'il y ait
branchement dans le cache. La sortie de cache est connectée au modèle de
langage par bloc, avec reprise de l'état atteint avant l'activation du cache.

Le cache contient en fait le dernier bloc du morceau de phrase en
cours, ce bloc peut être répété. Par contre, si c'est le bloc avant le dernier, il
20 ne peut être traité par un tel cache, il faut alors revoir entièrement la phrase.

Lorsqu'il s'agit d'une répétition sur des articles, et pour les langues
ou cela est pertinent, le cache comporte l'article et ses formes associées, en
changement de nombre et de genre.

25 En français par exemple, le cache pour "de" contient "du" et
"des". La modification de genre et de nombre est en effet fréquente.

Echange simple et changement d'avis :

30

L'échange simple est traité par la création de groupes de blocs
associés entre lesquels un échange simple est possible, c'est-à-dire qu'il

existe une probabilité pour qu'il y ait sortie du bloc et branchement au début d'un des autres blocs du groupe.

5 Pour l'échange simple, la sortie de bloc est couplée avec un déclenchement, dans les blocs associés au même groupe, de sous parties de même sens.

Pour le changement d'avis, soit il n'y a pas de déclenchement, soit il y a déclenchement sur les sous parties de sens distinct.

10

Il est également possible de n'avoir pas recours au déclenchement, et de classer l'hésitation par analyse à posteriori.

15

Bafouillage :

Il est traité comme une répétition simple.

20 L'avantage de ce mode de traitement des hésitations (sauf pour l'hésitation simple) est que la création des groupes associés renforce le taux de reconnaissance par rapport à une phrase sans hésitation, à cause de la redondance d'information sémantique présente. Par contre, la charge de calcul est plus importante.

25

Références :

- 5 (1) Self-Organized language modeling for speech recognition. F.Jelinek.
 Readings in speech recognition. P 450-506, Morgan Kaufman Publishers,
 1990
- (2) Basic methods of probabilistic context free grammars. F. Jelinek, J.D.
 Lafferty, R.L. Mercer NATO ASI Series Vol 75 P 345-359, 1992
- 10 (3) Trigger-Based language models : A maximum entropy approach. R.
 Lau, R. Rosenfeld, S. Roukos. Proceedings IEEE ICASSP, 1993
- (4) Statistical methods for speech recognition F. Jelinek MIT Press ISBN
 0-262-10066-5 pp 245-253

Revendications

1. Dispositif de reconnaissance vocale (1) comportant un
5 processeur audio (2) pour l'acquisition d'un signal audio, et un décodeur
linguistique (6) pour déterminer une séquence de mots correspondant au
signal audio, le décodeur comportant un modèle de langage (8), caractérisé
en ce que le modèle de langage (8) est déterminé par un premier ensemble
d'au moins un bloc syntaxique rigide et un second ensemble d'au moins un
10 bloc syntaxique souple.

2. Dispositif selon la revendication 1, caractérisé en ce que le
premier ensemble d'au moins un bloc syntaxique rigide est défini par une
grammaire de type BNF.

15

3. Dispositif selon les revendications 1 ou 2, caractérisé en ce que
le second ensemble d'au moins un bloc syntaxique souple est défini par un
ou plusieurs réseaux n-gram, les données des réseaux n-gram étant
produites à l'aide d'une grammaire ou d'une liste de tronçons de phrase.

20

4. Dispositif selon la revendication 3, caractérisé en ce que le
réseau n-gram contient des données correspondant à un ou plusieurs des
phénomènes suivants : l'hésitation simple, la répétition simple, l'échange
simple, le changement d'avis, le bafouillage.

25

1 / 1

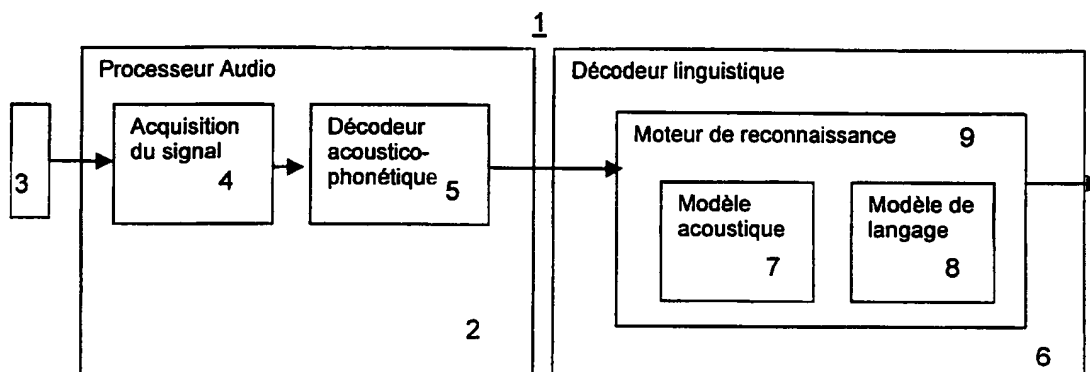


Fig. 1

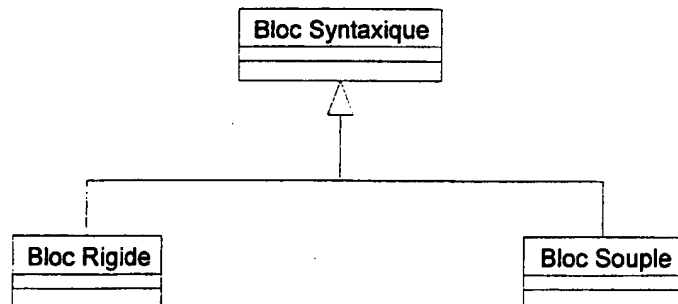


Fig. 2

INTERNATIONAL SEARCH REPORT

International Application No

PCT/FR 00/03329

A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 G10L15/18

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 G10L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, INSPEC, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	TSUKADA H ET AL: "Reliable utterance segment recognition by integrating a grammar with statistical language constraints" SPEECH COMMUNICATION, ELSEVIER SCIENCE PUBLISHERS, AMSTERDAM, NL, vol. 26, no. 4, December 1998 (1998-12), pages 299-309, XP004153051 ISSN: 0167-6393	1,3
Y	the whole document --- -/--	2,4

☒ Further documents are listed in the continuation of box C.

☐ Patent family members are listed in annex.

* Special categories of cited documents:

A document defining the general state of the art which is not considered to be of particular relevance

E earlier document but published on or after the international filing date

L document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

O document referring to an oral disclosure, use, exhibition or other means

P document published prior to the international filing date but later than the priority date claimed

T later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

X document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

Y document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

& document member of the same patent family

Date of the actual completion of the international search

27 February 2001

Date of mailing of the international search report

09/03/2001

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Quélavoine, R

INTERNATIONAL SEARCH REPORT

International Application No

PCT/FR 00/03329

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	LLOYD-THOMAS H ET AL: "AN INTEGRATED GRAMMAR/BIGRAM LANGUAGE MODEL USING PATH SCORES" DETROIT, MAY 9 - 12, 1995. SPEECH, NEW YORK, IEEE, US, 9 May 1995 (1995-05-09), pages 173-176, XP000657958 ISBN: 0-7803-2432-3 the whole document	1,3
Y	ROSE R C ET AL: "MODELING DISFLUENCY AND BACKGROUND EVENTS IN ASR FOR A NATURAL LANGUAGE UNDERSTANDING TASK" PHOENIX, AZ, MARCH 15 - 19, 1999, NEW YORK, NY: IEEE, US, 15 March 1999 (1999-03-15), pages 341-344, XP000900128 ISBN: 0-7803-5042-1 the whole document	4
Y	DATABASE INSPEC 'Online! INSTITUTE OF ELECTRICAL ENGINEERS, STEVENAGE, GB; CSABAY K: "Catalogue cards, UDC number, dictionary entries" Database accession no. 6231820 XP002161480 abstract & TUDOMANYOS ES MUSZAKI TAJEKOZTATAS, JAN. 1999, OMIKK, HUNGARY, vol. 46, no. 1, pages 12-32, ISSN: 0041-3917	2
A	AMENGUAL J -C ET AL: "Simplifying language through error-correcting decoding" PROCEEDINGS ICSLP 96. FOURTH INTERNATIONAL CONFERENCE ON SPOKEN LANGUAGE PROCESSING (CAT. NO.96TH8206), PROCEEDING OF FOURTH INTERNATIONAL CONFERENCE ON SPOKEN LANGUAGE PROCESSING. ICSLP '96, PHILADELPHIA, PA, USA, 3-6 OCT. 1996, pages 841-844 vol.2, XP002161479 1996, New York, NY, USA, IEEE, USA ISBN: 0-7803-3555-4 the whole document	4

-/--

INTERNATIONAL SEARCH REPORT

International Application No

PCT/FR 00/03329

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>METEER M ET AL: "STATISTICAL LANGUAGE MODELING COMBINING N-GRAM AND CONTEXT-FREE GRAMMERS"</p> <p>MINNEAPOLIS, APR. 27 - 30, 1993, NEW YORK, IEEE, US,</p> <p>vol. -, 27 April 1993 (1993-04-27), pages II-37-40, XP000427719</p> <p>ISBN: 0-7803-0946-4</p> <p>the whole document</p> <p>-----</p>	1-3

RAPPORT DE RECHERCHE INTERNATIONALE

De de internationale No

PCT/FR 00/03329

A. CLASSEMENT DE L'OBJET DE LA DEMANDE
CIB 7 G10L15/18

Selon la classification internationale des brevets (CIB) ou à la fois selon la classification nationale et la CIB

B. DOMAINES SUR LESQUELS LA RECHERCHE A PORTE

Documentation minimale consultée (système de classification suivi des symboles de classement)

CIB 7 G10L

Documentation consultée autre que la documentation minimale dans la mesure où ces documents relèvent des domaines sur lesquels a porté la recherche

Base de données électronique consultée au cours de la recherche internationale (nom de la base de données, et si réalisable, termes de recherche utilisés)

EPO-Internal, INSPEC, WPI Data

C. DOCUMENTS CONSIDERES COMME PERTINENTS

Catégorie *	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
X	TSUKADA H ET AL: "Reliable utterance segment recognition by integrating a grammar with statistical language constraints" SPEECH COMMUNICATION, ELSEVIER SCIENCE PUBLISHERS, AMSTERDAM, NL, vol. 26, no. 4, décembre 1998 (1998-12), pages 299-309, XP004153051 ISSN: 0167-6393	1,3
Y	le document en entier --- -/--	2,4



Voir la suite du cadre C pour la fin de la liste des documents



Les documents de familles de brevets sont indiqués en annexe

* Catégories spéciales de documents cités:

- *A* document définissant l'état général de la technique, non considéré comme particulièrement pertinent
- *E* document antérieur, mais publié à la date de dépôt international ou après cette date
- *L* document pouvant jeter un doute sur une revendication de priorité ou cité pour déterminer la date de publication d'une autre citation ou pour une raison spéciale (telle qu'indiquée)
- *O* document se référant à une divulgation orale, à un usage, à une exposition ou tous autres moyens
- *P* document publié avant la date de dépôt international, mais postérieurement à la date de priorité revendiquée

- *T* document ultérieur publié après la date de dépôt international ou la date de priorité et n'appartenant pas à l'état de la technique pertinent, mais cité pour comprendre le principe ou la théorie constituant la base de l'invention
- *X* document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme nouvelle ou comme impliquant une activité inventive par rapport au document considéré isolément
- *Y* document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme impliquant une activité inventive lorsque le document est associé à un ou plusieurs autres documents de même nature, cette combinaison étant évidente pour une personne du métier
- *Z* document qui fait partie de la même famille de brevets

Date à laquelle la recherche internationale a été effectivement achevée

27 février 2001

Date d'expédition du présent rapport de recherche internationale

09/03/2001

Nom et adresse postale de l'administration chargée de la recherche internationale
Office Européen des Brevets, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax (+31-70) 340-3016

Fonctionnaire autorisé

Quélavoine, R

RAPPORT DE RECHERCHE INTERNATIONALE

De : de Internationale No

PCT/FR 00/03329

C.(suite) DOCUMENTS CONSIDERES COMME PERTINENTS		
Catégorie	Identification des documents cités, avec, le cas échéant, l'Indication des passages pertinents	no. des revendications visées
X	<p>LLOYD-THOMAS H ET AL: "AN INTEGRATED GRAMMAR/BIGRAM LANGUAGE MODEL USING PATH SCORES"</p> <p>DETROIT, MAY 9 - 12, 1995. SPEECH, NEW YORK, IEEE, US,</p> <p>9 mai 1995 (1995-05-09), pages 173-176, XP000657958</p> <p>ISBN: 0-7803-2432-3</p> <p>le document en entier</p>	1, 3
Y	<p>ROSE R C ET AL: "MODELING DISFLUENCY AND BACKGROUND EVENTS IN ASR FOR A NATURAL LANGUAGE UNDERSTANDING TASK"</p> <p>PHOENIX, AZ, MARCH 15 - 19, 1999, NEW YORK, NY: IEEE, US,</p> <p>15 mars 1999 (1999-03-15), pages 341-344, XP000900128</p> <p>ISBN: 0-7803-5042-1</p> <p>le document en entier</p>	4
Y	<p>DATABASE INSPEC 'en ligne!</p> <p>INSTITUTE OF ELECTRICAL ENGINEERS, STEVENAGE, GB;</p> <p>CSABAY K: "Catalogue cards, UDC number, dictionary entries"</p> <p>Database accession no. 6231820</p> <p>XP002161480</p> <p>abrégé</p> <p>& TUDOMANYOS ES MUSZAKI TAJEKOZTATAS, JAN. 1999, OMIKK, HUNGARY,</p> <p>vol. 46, no. 1, pages 12-32,</p> <p>ISSN: 0041-3917</p>	2
A	<p>AMENGUAL J -C ET AL: "Simplifying language through error-correcting decoding"</p> <p>PROCEEDINGS ICSLP 96. FOURTH INTERNATIONAL CONFERENCE ON SPOKEN LANGUAGE PROCESSING (CAT. NO.96TH8206), PROCEEDING OF FOURTH INTERNATIONAL CONFERENCE ON SPOKEN LANGUAGE PROCESSING. ICSLP '96,</p> <p>PHILADELPHIA, PA, USA, 3-6 OCT. 1996,</p> <p>pages 841-844 vol.2, XP002161479</p> <p>1996, New York, NY, USA, IEEE, USA</p> <p>ISBN: 0-7803-3555-4</p> <p>le document en entier</p>	4

-/--

RAPPORT DE RECHERCHE INTERNATIONALE

De ... de Internationale No

PCT/FR 00/03329

C.(suite) DOCUMENTS CONSIDERES COMME PERTINENTS

Catégorie	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
A	<p>METEER M ET AL: "STATISTICAL LANGUAGE MODELING COMBINING N-GRAM AND CONTEXT-FREE GRAMMERS"</p> <p>MINNEAPOLIS, APR. 27 - 30, 1993, NEW YORK, IEEE, US,</p> <p>vol. -, 27 avril 1993 (1993-04-27), pages II-37-40, XP000427719</p> <p>ISBN: 0-7803-0946-4</p> <p>le document en entier</p> <p>-----</p>	1-3